General Description & Theoretical Background

Stage theories assume that behavior change involves movement through a sequence of discrete stages, that different variables influence different stage transitions, and that effective interventions need to be matched to stage (Sutton, 2005; Weinstein, Rothman & Sutton, 1998). Stage is the central construct in the transtheoretical model (TTM; Prochaska & Velicer, 1997), the precaution adoption process model (PAPM; Weinstein, 1988; Weinstein & Sandman, 1992), the stage version of the health action process approach (HAPA; Schwarzer, 2008), the health behavior goal model (Maes & Gebhardt, 2000), the Rubicon model of action phases (Heckhausen, 1991), the perspectives on change model of smoking cessation (Borland, Balmford & Hunt, 2004), the AIDS risk reduction model (Catania, Kegeles & Coates, 1990), and theories of delay in seeking health care (e.g., Andersen, Cacioppo & Roberts, 1995).

Different stage theories use different definitions and measures of stages. This article uses the TTM and the PAPM to illustrate the construct of stage and how it is measured. First, the two theories will be briefly described.

The Transtheoretical Model

The TTM is the dominant stage model in health psychology and health promotion. It was developed in the 1980s by James Prochaska and colleagues at the University of Rhode Island. The model includes several constructs, with the stages of change providing the basic organizing principle. The most widely used version of the model specifies five stages: precontemplation, contemplation, preparation, action and maintenance. People are assumed to move through the stages in order, but they may regress from action or maintenance to an earlier stage. People may cycle through the stages several times before achieving long-term behavior change. The model has been applied to a wide range of health behaviors, with smoking cessation still the most frequent application.

The TTM has been very influential and has popularized the idea that behavior change involves movement through a series of stages. It has also stimulated the development of innovative interventions. However, the TTM has attracted a large amount of criticism over the years, culminating in a recent call for the model to be abandoned (West, 2005). The main problems with the model concern the definition and measurement of the stages (discussed below) and the lack of a clear specification of which variables influence which stage transitions. In addition, most supportive research has relied on cross-sectional designs rather than stronger...
research designs, in particular longitudinal studies of stage transitions and experimental studies of matched and mismatched interventions (Sutton, 2005; Weinstein et al., 1998).

The Precaution Adoption Process Model

The PAPM was originally developed to describe and explain the process by which people adopt precautions against a new risk, i.e., one that they have recently learned about rather than one they have been aware of for some time. For example, the model was applied to understanding the adoption of precautionary behavior after warnings were released about the high levels of radon in homes in specific geographic areas. It is also applicable in the situation where a new precaution against an “old” risk becomes available (e.g., the introduction of the HPV vaccination to prevent cervical cancer.

The model specifies seven discrete stages. In Stage 1, people are unaware of the health issue. People in Stage 2 are aware of the issue but they have never thought about adopting the precaution; they are not personally engaged by the issue. People who reach Stage 3 are personally engaged but they are undecided about whether to adopt the precaution. If they decide against adopting the precaution, they move into Stage 4, or out of the sequence of action adoption. If they decide in favor, but have not yet acted on this decision, they are in Stage 5. People who act on their decision move to Stage 6. Finally, for some behaviors, a seventh stage (maintenance) may be appropriate.

Although only a handful of studies using the PAPM have been conducted to date, it is a promising approach that avoids some of the problems with the TTM. For example, it defines the stages without reference to arbitrary time periods and, e.g., between having never thought about adopting a particular precaution and having thought about it and decided not to act. Key tasks for future research on the PAPM are to specify the variables that are important for each of the stage transitions and to test whether they predict and influence these transitions.

Measuring Stage

The Transtheoretical Model

In work on the TTM, three main methods have been used to measure stages: multi-dimensional questionnaires; single-item continuous measures of readiness to change; and staging algorithms and self-categorizations. These are discussed in turn.

Multi-dimensional questionnaires. Most studies that have investigated alcohol and drug use from the standpoint of the TTM have used multi-dimensional questionnaires to measure stage of change. In this approach, each stage is measured by a set of questionnaire items, and scores are derived for each individual representing their position on each dimension. Three such multi-dimensional questionnaires have been used in studies of alcohol and drug use: the University of Rhode Island Change Assessment (URICA; McConnaughey, Prochaska & Velicer, 1983),
the Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES; Miller & Tonigan, 1996) and the Readiness to Change Questionnaire (RTCQ; Rollnick, Heather, Gold & Hall, 1992). Consider the URICA as an example of this approach. The URICA was the first multi-dimensional questionnaire designed to measure stages of change, and is intended for use in clinical contexts. It consists of 32 items, eight for each of four stages (precontemplation, contemplation, action, maintenance; Appendix A). The items refer generically to the respondent’s “problem” but do not specify a particular problem behavior.

In studies using the URICA, there is a fairly consistent pattern of relatively large correlations among subscales representing adjacent stages. The correlations between contemplation and maintenance, which are non-adjacent stages, are also relatively large (Sutton, 2001). The observed pattern of correlations suggests that the URICA is not measuring discrete stages. The same applies to the SOCRATES and the RTCQ. Thus, multi-dimensional questionnaires are not consistent with the assumption of discrete stages.

**Single-item continuous measures of readiness to change.** Biener and Abrams (1991) developed the Contemplation Ladder as a simple measure of readiness to consider smoking cessation. This scale consists of a visual representation of a vertical ladder scaled from 0 to 10, with the scalepoints 0, 2, 5, 8 and 10 labeled as follows: “No thought of quitting”/”Think I need to consider quitting someday”/”Think I should quit but not quite ready”/”Starting to think about how to change my smoking patterns”/”Taking action to quit (e.g., cutting down, enrolling in a program)”. In their sample of smokers recruited from two worksites, Biener and Abrams (1991) found a highly significant correlation of 0.64 between the ladder score and a single-item measures of intention to try to quit. The ladder and intention measures showed similar patterns of correlations with other variables. With regard to predictive validity, ladder scores were more predictive than were intention scores of participation in cotinine assessment and taking a self-test of addiction to nicotine, which Biener and Abrams interpret as earlier stages of readiness, whereas intention was a better predictor of participation in events that required a quit attempt and of successful quitting. Note that the contemplation ladder does not represent the full set of stages: it does not include the action stage (usually defined in work on the TTM as having stopped smoking for up to 6 months) or the maintenance stage (having quit for more than 6 months). The contemplation ladder has been adapted for use with other behaviors, for example marijuana use (Slavet et al., 2006) and gambling (Petry, 2005).

Similar measures were developed by LaBrie, Quinlan, Schiffman and Earleywine (2005) for alcohol and safer sex. These readiness-to-change “rulers” consist of a horizontal line scaled from 0 to 10 with five labels (e.g., for drinking: “Never think about my drinking”/”Sometimes I think about drinking less”/”I have decided to drink less”/”I am already trying to cut back on my drinking”/”My drinking has changed. I now drink less than before”). The alcohol change ruler correlated 0.77 with total score from the RTCQ (Rollnick et al., 1992). Similarly, the safer sex ruler correlated 0.77 with the total score from an 11-item Readiness to Change Risky Sexual Behavior (RTCQ-SB) based on the RTCQ. The rulers showed higher correlations with measures of intentions to change
than did the scores from the corresponding multi-item scales. The authors conclude that the rulers may be useful when a quick and simple assessment of readiness to change is required.

Virtually all the studies that have used contemplation ladders or readiness rulers have treated the scores as continuous measures of readiness to change. This approach is inconsistent with the idea that change involves movement through a sequence of discrete stages. Furthermore, the use of measures of behavioral intention as criteria for assessing the validity of ladders and rulers raises the question of whether intention measures could themselves be used as a simple way of assessing readiness to change.

Staging algorithms and self-categorizations. A staging algorithm uses a small number of questionnaire items and a set of rules to allocate participants to stages in such a way that no individual can be in more than one stage. Self-categorizations are single-item measures in which participants are presented with a list of statements, each of which represents a stage, and are asked to select the one that best describes them. Both these methods are relatively brief and simple, compared to multi-dimensional questionnaires. (Staging algorithms could be highly complex but in practice they use a small number of items and a simple set of rules). If the resulting measure is analysed as a set of categories rather than as a continuous scale, this approach is in principle more consistent with the assumption of discrete stages than is either of the other two methods.

The few studies that have compared staging algorithms with multi-dimensional questionnaires have found low concordance between them (e.g., Belding, Iguchi & Lamb, 1996), suggesting that they are measuring quite different constructs. A study that compared a staging algorithm with the contemplation ladder divided into three categories (analogous to “precontemplation”, “contemplation” and “preparation”) found a correlation of 0.58 between the two classification schemes but also some important differences (Herzog, Abrams, Emmons & Linnan, 2000). For example, of those who were classified by the algorithm as precontemplators, 49% were in the “contemplation” group as assessed by the ladder, suggesting that the two measures should not be used interchangeably.

A staging algorithm for smoking that has been used in a large number of studies since it was introduced by DiClemente and colleagues (DiClemente, Prochaska, Fairhurst, Velicer, Velasquez & Rossi, 1991) can be found in Appendix B. Precontemplation, contemplation, and preparation are defined in terms of current behavior, intentions and past behavior (whether or not the person has made a 24-hour quit attempt in the past year), whereas action and maintenance are defined purely in terms of behavior; ex-smokers’ intentions are not taken into account. Although the phrase “seriously thinking of quitting” is used in Appendix B, different versions of this algorithm have used alternative wordings including “seriously considering quitting”, “intending to quit” and “planning to quit”. Such apparently minor changes in wording can have a large effect on stage distributions (Sutton, 2000).
Critics have pointed out a number of problems with this algorithm, some of which stem from the way that contemplation and preparation are defined (e.g., Borland, Balmford, Segan, Livingston & Owen, 2003; Etter & Sutton, 2002). According to this algorithm, a smoker cannot be in the preparation stage unless he or she has made a recent quit attempt. Thus, a smoker can never be "prepared" for his or her first quit attempt. Similarly, the subgroup of smokers in the contemplation stage who intend to quit in the next 30 days but have not made a quit attempt in the past year cannot move to the preparation stage. Thus, the stages are defined in such a way that some smokers cannot move directly to the next stage in the sequence (Sutton, 2000).

A variant of the TTM developed by a group of researchers in the Netherlands (e.g., Dijkstra, Bakker & De Vries, 1997) uses different definitions of the stages that avoid these problems. In this model, the pre-action stages are defined purely in terms of intention: preparation is defined as planning to quit in the next month and contemplation as planning to quit in the next six months but not in the next month.

A staging algorithm for exercise developed by Marcus and Simpkin (1993) is in Appendix C. Although this algorithm does not suffer from the logical problems of the DiClemente et al. (1991) smoking algorithm, it seems somewhat implausible to treat irregular exercise (preparation) as a discrete stage between contemplation and action, implying that people move from no exercise to irregular exercise to regular exercise and that irregular exercise is qualitatively different from regular exercise.

A problem affecting many TTM staging algorithms is that the time periods are arbitrary. For instance, action and maintenance are usually distinguished by whether or not the duration of behavior change exceeds six months. Changing the time periods would lead to different stage distributions. The use of arbitrary time periods casts doubt on the assumption that the stages are qualitatively distinct, in other words that they are true stages rather than “pseudostages” – arbitrarily created segments of an underlying continuous variable (Bandura, 1997).

Comparison of the algorithms in Appendices B and C illustrates another problem: TTM staging algorithms for different health behaviors often use inconsistent stage definitions. For example, in the algorithm for adoption of mammography (Rakowski, Dube, Marcus, Prochaska, Velicer & Abrams, 1992), action and maintenance are defined partly in terms of intentions (planning to have a mammogram in the coming year). It is possible for a woman to move directly from contemplation to maintenance simply by forming an intention, without passing through the action stage and without changing her behavior.

The Precaution Adoption Process Model

Unlike the TTM, in which a variety of measurement methods have arisen, only staging algorithms have been used in research on the PAPM to date. Appendix D gives a stage classification algorithm that would be suitable for any behavior for which a maintenance stage is not applicable (Weinstein & Sandman, 2002). These include behaviors that, if they are performed at all, are usually performed only once,
e.g., having a predictive genetic test for inherited breast/ovarian cancer. Of course, virtually any behavior can be repeated: persons may test their home for radon and then retest it two years later. If a significant proportion of people in the sample have adopted the precaution before, then it may be necessary to take past behavior into account in the analysis and to reword the staging algorithm. Consider, for example, applying the model to participation in mammography screening. If the investigator is interested in first-time attendance for screening, he or she could either select a sample of women who have recently reached the lower age limit for screening and use the algorithm in Appendix D to stage them or select a sample of women who have never been screened and follow them over time until some of them have their first screen, using the algorithm to stage the sample on a number of occasions. Women who have had one mammogram could be allocated to stages with respect to having another mammogram. This would require modifications to the algorithm. Stage 1 would not be applicable for these women. And the statement used to classify women in Stage 2 could be reworded to something like “I haven’t thought about whether to have another mammogram”. (An alternative approach would be to classify women who have had repeated mammograms in accordance with the recommended schedule as being in the maintenance stage. However, it would be difficult to know how to classify women who have had more than one mammogram but whose pattern of attendance does not conform to the recommended schedule.)

For ongoing behaviors such as the frequency of exercising or the amount of salt consumed per day, it is necessary to define a criterion level of behavior, for example doing at least 30 minutes of moderate physical activity every day. In this case, it would be appropriate to specify a maintenance stage, possibly defined in terms of duration as in the TTM, for example having maintained at least 30 minutes of moderate physical activity a day for at least six months. However, as noted above, such time periods are arbitrary and do not have face validity as marking a transition between discrete stages.

Reliability and Validity of Staging Algorithms and Self-categorizations

Although a staging algorithm uses several questionnaire items, it yields a single measure of stage. Similarly, self-categorizations are single-item measures. Indices of internal consistency such as Cronbach’s alpha are therefore not applicable. However, test-retest reliability can be assessed, by measuring stage on two occasions in the same sample. A short time interval (e.g., one week) should be used to reduce the likelihood that true change occurs. Several studies have assessed test-retest reliability of TTM staging algorithms over short time periods (e.g., Aveyard, Lancashire, Almond & Cheng, 2002; Carey, Purnine, Maisto & Carey, 2002; Donovan, Jones, Holman & Corti, 1998; Etter & Sutton, 2002), yielding kappas that ranged from 0.46 to 0.83 (“moderate” to “very good” agreement according to Altman, 1991). The test-retest reliability of the component questions can also be assessed (e.g., Aveyard et al., 2002). The reliability of the PAPM algorithm has not been assessed to date.

There are several ways of assessing the validity of a staging algorithm or self-categorization. First, these measures incorporate a measure of the behavior (whether the individual is in a pre-action or a post-action stage). Sometimes it may be
possible to assess the validity of this aspect of the stage measure against another measure of the behavior that can be assumed to be more valid. In the case of smoking, for example, a biochemical measure such as saliva cotinine concentration could be used to verify self-reported smoking status (i.e., pre-action versus post-action). Use of this method requires that the algorithm incorporates a clear and precise definition of the behavior. Furthermore, this method provides no evidence bearing on the validity or otherwise of the distinction between the pre-action stages or of the distinction between the post-action stages.

A second way to estimate the validity of the component items of a staging algorithm depends on the availability of suitable criteria. For example, intention to obtain a mammogram in the next 6 months can be validated against screening office records of attendance. Finally, the construct validity of a staging algorithm or self-categorization can be assessed by examining whether the stage measure is related to measures of the other constructs in the theory in ways that are predicted by the theory. Thus, a test of the construct validity of a stage measure is inherent in testing predictions from a stage theory. Because there is no ideal exemplar study in the literature, a hypothetical example is given: If a stage theory makes a particular prediction involving the stages, for example, that variables a, b and c influence the transition from stage I to stage II and variables c and d influence the transition from stage II to stage III, support for this prediction in a longitudinal study of stage transitions also provides support for the validity of the stage measure (and for the other measures). However, if the prediction is not supported, the problem could lie with the theory (it is misspecified in this respect) or with the validity of the stage measure or the measures of the other constructs.

Staging algorithms and self-categorizations differ in the degree to which they are transparent to the respondent. With a self-categorization measure, the respondent can see the full set of stages before selecting the one that best applies to them. By contrast, with a staging algorithm, the rules governing allocation to stages are usually not available to the respondent. This difference may have implications for reliability and validity. For example, compared with staging algorithms, self-categorizations may be easier to understand but more susceptible to social desirability bias.

**Recommendations**

1. **Staging algorithms and self-categorizations are conceptually more consistent with the assumption of discrete stages than are multi-dimensional questionnaires and continuous measures such as contemplation ladders and readiness rulers.** Although staging algorithms and self-categorizations may be less reliable than multi-dimensional questionnaires, they should be the method of choice for researchers and practitioners who test stage theories.

2. **Researchers who plan to use the TTM may wish to use the algorithms developed by the Rhode Island group for comparability with previous research. However, they should also consider including alternative stage algorithms that avoid some of the problems identified with the standard measures.**
3. In developing new stage algorithms or self-categorizations, **pre-action stages should not be defined in terms of past behavior so as to avoid logical problems.** **Arbitrary time periods should also be avoided.** Stages should be defined in terms of events, for example making a decision to change one’s behavior or setting a quit date. This may involve modifying the original theory or developing a new theory, so that the stage measures accurately reflect the theoretical stage definitions.

4. **Stage measures should incorporate a clear and precise definition of the behavior.**

5. In testing predictions from stage theories (and thus at the same time testing the construct validity of stage measures), **researchers should use strong research designs, in particular longitudinal prediction of stage transitions and experimental studies of matched and mismatched interventions.**
References


Appendix A. Example items from the University of Rhode Island Change Assessment (URICA)

Precontemplation
“As far as I’m concerned, I don’t have any problems that need changing”
“All this talk about psychology is boring. Why can’t people just forget about their problems?”

Contemplation
“I have a problem and I really think I should work on it”
“I’m hoping this place will help me to better understand myself”

Action
“I am doing something about the problems that had been bothering me”
“Anyone can talk about change: I’m actually doing something about it”

Maintenance
“It worries me that I might slip back on a problem I have already changed, so I am here to seek help”
“I’m here to prevent myself from having a relapse of my problem”
Appendix B. TTM staging algorithm for adult smoking, from University of Rhode Island Cancer Prevention Research Center (2008)

Are you currently a smoker?
- Yes, I currently smoke
- No, I quit within the last 6 months (ACTION STAGE)
- No, I quit more than 6 months ago (MAINTENANCE STAGE)
- No, I have never smoked (NONSMOKER)

(For smokers only) In the last year, how many times have you quit smoking for at least 24 hours?

(For smokers only) Are you seriously thinking of quitting smoking?
- Yes, within the next 30 days (PREPARATION STAGE if they have one 24-hour quit attempt in the past year – refer to previous question… if no attempt then CONTEMPLATION STAGE)
- Yes, within the next 6 months (CONTEMPLATION STAGE)
- No, not thinking of quitting (PRECONTEMPLATION STAGE)
Appendix C. Staging algorithm for exercise, from Marcus & Simpkin (1993)

<table>
<thead>
<tr>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I currently do not exercise</td>
</tr>
<tr>
<td>2. I intend to exercise in the next 6 months</td>
</tr>
<tr>
<td>3. I currently exercise regularly</td>
</tr>
<tr>
<td>4. I have exercised regularly for the past 6 months</td>
</tr>
</tbody>
</table>

**Scoring**

Precontemplation: Item 1 = true and Item 2 = false
Contemplation: Item 1 = true and Item 2 = true
Preparation: Item 1 = false and Item 3 = false
Action: Item 3 = true and Item 4 = false
Maintenance: Item 3 = true and Item 4 = true
Appendix D. PAPM stage classification algorithm, from Weinstein & Sandman (2002)

1. Have you ever heard about {home radon testing}?
   - No                                             Stage 1
   - Yes [go to 2]

2. Have you {tested your own house for radon}?
   - Yes                                             Stage 6
   - No [go to 3]

3. Which of the following best describes your thoughts about {testing your home}?
   - I've never thought about {testing}               Stage 2
   - I'm undecided about {testing}                    Stage 3
   - I've decided I don't want to {test}              Stage 4
   - I've decided I do want to {test}                 Stage 5

Note: The words in curly brackets could be replaced with other precautions to develop a staging algorithm for these precautions.

Key to stages:
- Stage 1: Unaware of issue
- Stage 2: Unengaged by issue
- Stage 3: Deciding about acting
- Stage 4: Decided not to act
- Stage 5: Decided to act
- Stage 6: Acted